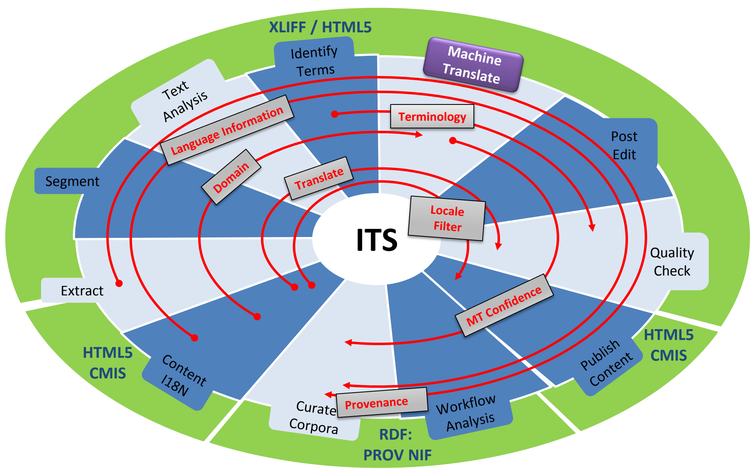
SIMPLE SEGMENT MACHINE TRANSLATION

Using DCU's MaTrEx for translating segments extracted from ITS 2.0-aware XLIFF / HTML Documents



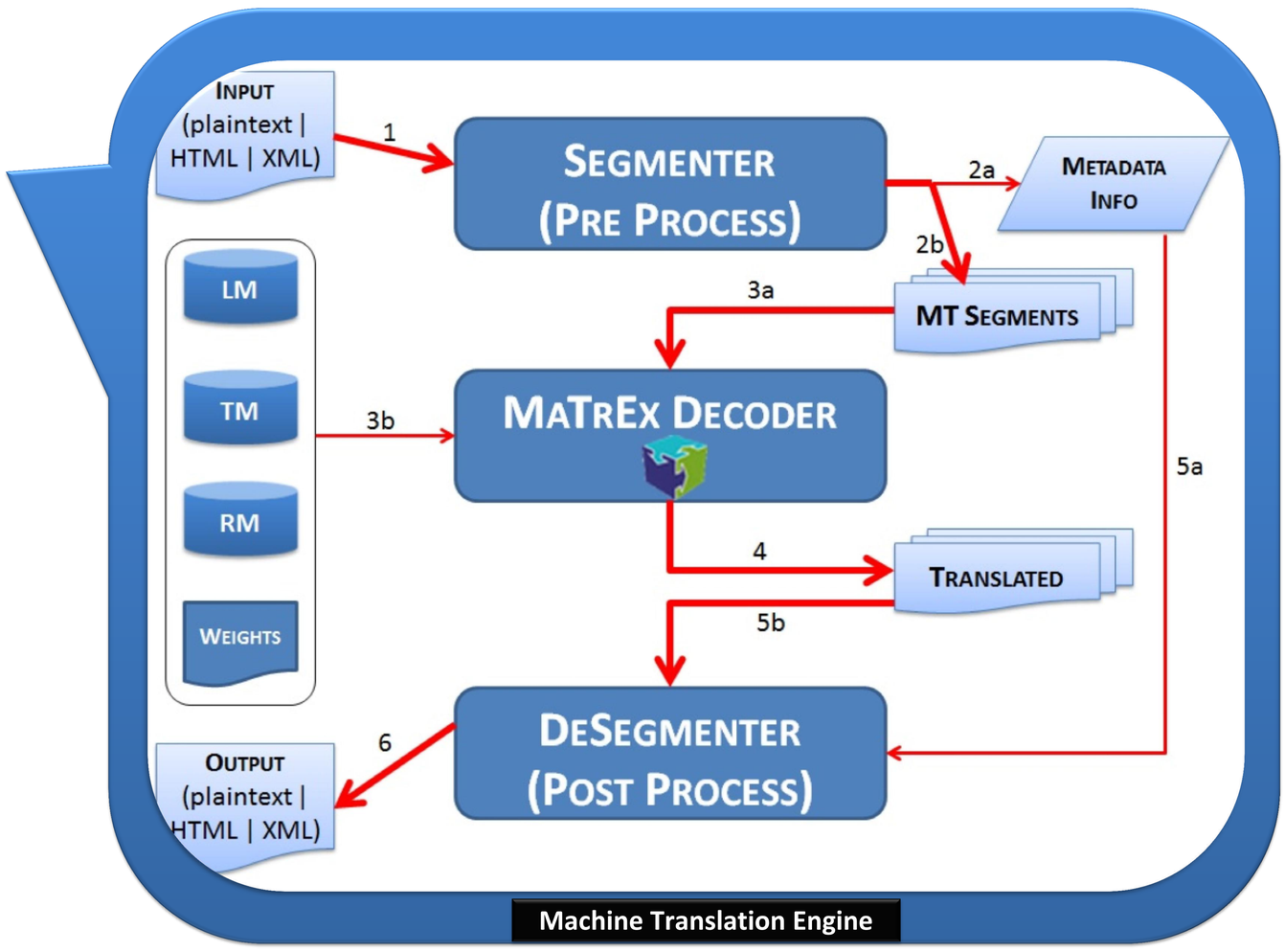
OBJECTIVES

- Demonstrate the translation of XLIFF (XML Localisation Interchange File Format) documents and HTML (HyperText Markup Language) webpages tagged with ITS (Internationalization Tag Set) 2.0
- Employ DCU's MaTrEx MT (Machine Translation) system as a Web Service – *a Statistical MT (SMT) system using the open-source Moses decoder interfaced with ITS 2.0 metadata-aware pre-processing and post-processing modules*
- Investigate the benefits of relevant ITS 2.0 data categories in the MT Workflow

Segmenter takes as input ITS 2.0 tagged document, parses it, and generates segments to be translated [Pre Process]

Decoder translates the segments with the help of statistical translation and language models

Desegmenter rebuilds the document by merging the translated segments with relevant ITS 2.0 metadata [Post Process]



BENEFITS

ITS 2.0 Data Category	Usage / Benefits to the System	Process / Stage
DOMAIN	Enables seamless application of domain-tuned MT engines	MT: Pre
LANGUAGE INFO	Enables easy identification of source / target language	MT: Pre / Post
LOCALE FILTER	Enables locale-specific translation operations	MT: Pre / Post
MT CONFIDENCE	Enables scores to be displayed accurately and automatically to PEs	MT: Dec / Post
PROVENANCE	Enables localisation workflow managers to compare performance	MT: Pre / Post
TERMINOLOGY	Enables terms to be identified for translations to be enforced in MT	MT: Decode
TRANSLATE	Ensures identification of text fragments that need to be preserved	MT: Pre / Dec

MaTrEx Web Service available at http://www.cngl.ie/mlwlt/

ITS2.0: The "Dublin Core" of the Multilingual Web

The MultilingualWeb-LT Working Group receives funding by the European Commission (project name LT-Web) through the Seventh Framework Programme (FP7) in the area of Language Technologies. Grant Agreement No. 287815.